Deep Learning based Facial Emotion Recognition

Swayam Badhe[a], Sangita Chaudhari
Department of Computer Engineering, Ramrao Adik Institute of Technology,
D. Y. Patil Deemed to be University, Nerul, Navi Mumbai India

Abstract. Human Facial expression is the mirror of human emotions playing very significant role in nonverbal communication. In many applications like human machine interface, medical diagnosis, AI based games, market research etc. needs facial expression recognition. Although it is very easy task for humans, its bit challenging for machines to detect and recognize correct emotions from series of human facial expressions. Since decade, many researchers have tried different image processing, machine learning and deep learning-based approaches to correctly identify human emotions. Some of them could able to identify emotions but with more complexity. WE have proposed CNN model with 4 convolution layer and 2 FC layers which is giving good accuracy over the existing models with less complexity. It is performing better for all classes except fear and disgust.

Keywords: Emotion Recognition, Facial Expressions, Emotion Classification, CNN

1. Introduction

In recent years, there is huge outbreak in Image processing and pattern recognition domain with the introduction of machine learning and deep learning approaches. It gives opportunities to researchers to avail automated feature extraction instead of hand-crafted feature extraction with better accuracy. Human emotion detection is always an important issue to be solved in many human computers interface-based application. Human visual system was so developed that it could easily detect and identify various emotions from facial expressions of human. However, to do so with machine is very challenging task and it needs perfect learning for machines to correctly classify the human emotions from facial expressions. It involves separation of various facial states from static human face images or from dynamic video sequence to determine the emotions of the object (face).

In today's era, there is huge demand of automated human emotion recognition in various applications like identification and access control, human machine interface, automated surveillance, medical diagnosis, AI based games etc. There are 7 classes of human emotions which can be identified and used in various applications to make out decisions. There emotions classes are Neutral, Angry, Disgust, fear, happy, sadness and surprise (figure 1). Each of these emotions results in changes in human facial components like lips, eyes, chin, cheeks, forehead, eyebrows. There orientation, colour, size, shapes may get changed. In automated human emotion recognition system, most of the approaches tries to extract features based on this and classify them in correct emotion. The objective of this work is to develop Automatic Facial Expression Recognition System which considers human face images/video and recognize and classify the emotions on human face into seven different expression classes as shown in figure 1.

1. Happy
2. Surprise
3. Neutral
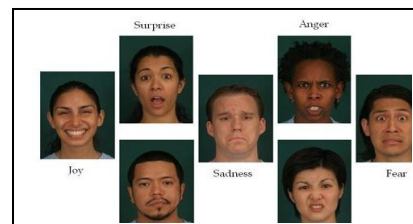4. Sad
5. Anger
6. Disgust
7. Fear

Fig 1. Basic Facial Expressions

Our goal will not only be to develop an Automatic Facial Expression Recognition System but also improving the accuracy of this system compared to the other available systems for various emotion classes.

2. Literature Survey

Over the years, many attempts have been made to detect and recognize emotions from facial expression as it has its own importance in human computer interaction used in variety of applications. Various Techniques like PCA, LDA, Curvelets, Gabor filters, SVM, AdaBoost, Local binary patterns, Histogram based descriptors, Neural network and its variants have been adopted in automated human emotion detection and recognition. Table II shows summary of comparison of few of the existing techniques based on PCA, SVM, AdaBoost, Local binary patterns and CNN [1-6].

[a] Electronic Mail: badheswayam@gmail.com

Table I. Comparison of Existing Systems

Paper	Technique Used	Accuracy (%)	Emotions considered	Limitations
Mutsugu et al. [1]	CNN Curvelet face +LDA Curvelet face+PCA	97.6 83.5 83.9	Happy, Neutral	Individuality of facial expression not detected
Zhang et al. [2]	Patch based 3D Gabor filters + SVM and AdaBoost	92.93	Happy, Neutral, Sadness, Surprise	JAFFE DB requires larger sizes of patches
Hayat [3]	SVM+ clustering	94.34	Anger, Disgust, Happiness, Fear, Sadness, Surprise	Less accuracy
Hablani [4]	Local binary patterns	94.44	Happy, Neutral, Sadness, Surprise, Anger, Fear, Disgust	Manual feature extraction, does not perform well for temporal behavior of facial expressions
Zisheng [5]	Pyramid HOG	96.33	Happy, Neutral, Sadness, Surprise and Anger, Fear and Disgust	less accuracy
Yu [6]	Deep CNN	61.2	Happiness, Disgust, Fear, Angry, Surprise and Sadness	Less accuracy, complex network

Mutsugu et al. used CNN with curvelet, PCA and LDA for emotion detection, but could recognize only Happy, Neural and lip movement capturing. The system proposed by them is insensitive to individuality of facial expressions [1]. In SVM based approaches, the classes considered are 7 but these approaches could not give good accuracy and it require larger database with larger patch size [2, 3]. In most of the approaches, emotions can be easily recognized from static facial images, but results out in poor accuracy in case of temporal changes in the facial expressions [4, 5]. In deep learning-based approach proposed by Yu, 7-layer CNN is used. Although majority of the emotion classes are considered, it does not give good accuracy at the cost of complexity in the used model. Similar approaches based on VGG, AlexNet and Resnet are also studied and it was found that although they tend to give moderate accuracy for temporal changes in human emotions, their complexity is very high [7-9].

3. Proposed Methodology

Human being is a gifted species with many emotions to have effective communications with humans and other species. Most common emotions emanated through human facial expressions are neutral, happy, sad, surprise, fear, anger and disgust. All these emotions are well expressed by contraction and expansion of various facial muscles. They are reflectance of human mind and changing dynamically with variety of thoughts in mind representing very complex signal with ample amount of information about our state of mind. One of the challenging tasks in artificial intelligence is recognize these emotions in various applications to make it more user friendly and to study the impact of contents and services on user. For example, customer's interest can be very well measured using facial recognition. Patients' facial expression can be monitored in response to the treatment and accordingly change in line of treatment can be followed in healthcare system. Human visual system is so developed that such expressions can easily identified and used in our day-to-day activities. However,

machine is not yet so intelligent to identify/recognize emotions and use it in decision making. There are many attempts made by researchers to devise solutions to recognize human emotions. We are proposing simple deep learning neural network model to make machine to recognize emotions and make inference out of it. Figure 2 shows general structure of facial recognition system.
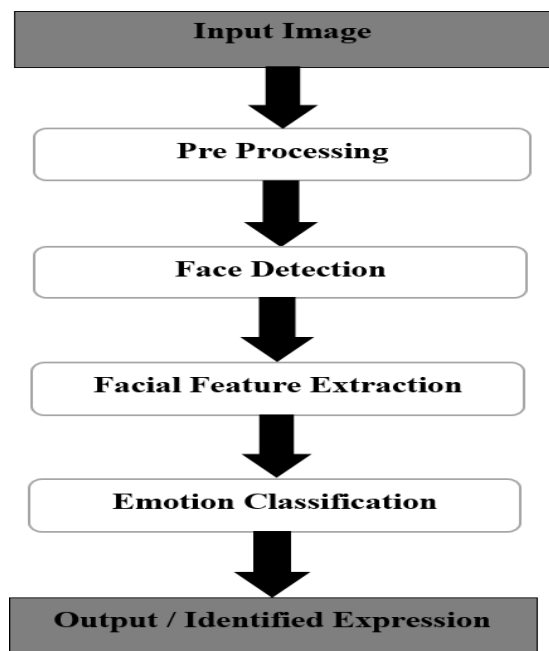

Fig. 2. Generalized Face Recognition System

In most of the systems, preprocessing, face detection, facial feature extractions and classification are common steps. Locating the face in the given scene is one of the crucial tasks in the crowded face scene. Once, face is detected, hand crafted facial features like shape, skin texture, face parts orientation, colour etc. are extracted and finally classification step will help to categories facial expression in one of the categories based on features. To improve the accuracy, deep convolutional neural network model is used in our proposed facial emotion recognition system. Figure 3 shows simple layer structure of convolution neural network.
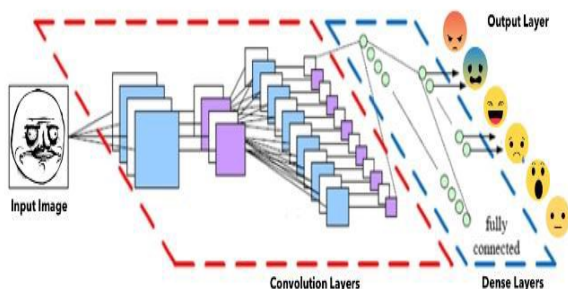

Fig. 3. Convolution Neural Network Layers

It is feedforward network with series of convolutional and pooling layers. Convolutional layer is performing the task of feature representation of the input image. Pooling

layer tries to reduce spatial resolution of feature maps with maximum value pooling from input feature map. Fully connected layer followed by these layers interprets these resultant feature maps to perform high level reasoning with SoftMax operator on top of it to achieve classification. Algorithm 1 shows the series of steps performed to recognize emotions from facial expressions.

Algorithm1: Facial Recognition
1. Resize input images Image to 48x48 pixels with labeling in to 7 classes (anger, disgust, fear, happiness, sadness, surprise, and neutral)
2. Apply preprocessing to remove noise and enhance images using median filtering and contrast enhancement.
3. Detect the face from each image
4. Detected face conversion to gray scale image
5. Pass the input image [1x48x48] to convolution2D layer.
6. Perform convolution to generate feature maps.
7. Apply Maxpooling2D using (2, 2) window across the calculated feature map by keeping maximum pixel value only.
8. Training with fully connected neural network with forward/backward propagation on the pixel values.
9. Use SoftMax function to represent probability of emotion class.

4. Results and Discussion

The proposed model is trained using 35887 images from FER2013 dataset. It consists of 48x48 grayscale face images. All images are pre-cropped with same region occupancy in image. The dataset consists of 8989 images for Happy, 6077 images for Sad, 6198 images for *Neutral,* 4002 images for *Surprise,* 5121 images for *Fear,* 4593 images for *Angry and* 547 images for *Disgust.* The facial images are to be classified in one of these seven classes: Happy, Sad, Neutral, Surprise, Fear, Angry and Disgust. Figure 4 shows sample images from FER2013 dataset.

Fig 4. FER2013 Emotion Data [10]

The model is performing really well for Happy, Sad, surprise, angry and neutral and able to predict the emotions of real time dynamic expression variation in face. Figure 5 shows visual results of Happy expression recognition with the prediction array. Similarly figure 6-9 shows visual results with the prediction array for Sad,

surprise, angry and neutral respectively. The proposed models' results are compared with the existing Model used in many other face recognition systems (Table II). It is observed that proposed model is giving good accuracy compared to other models for 5 classes and results out in less complexity as well. It could classify all emotions correctly except fear and disgust misclassified occasionally.
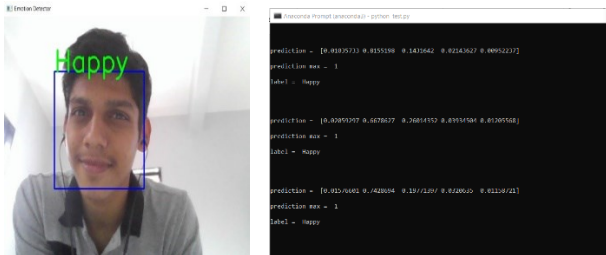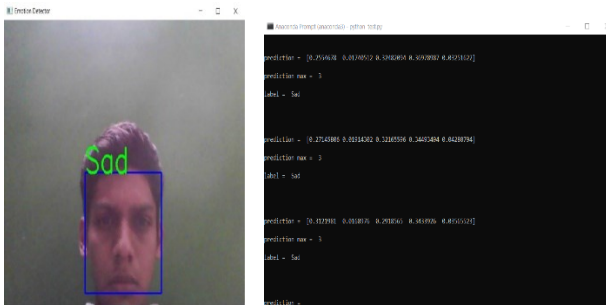

Fig 5. Prediction of Happy Emotion

Fig 6. Prediction of Sad Emotion
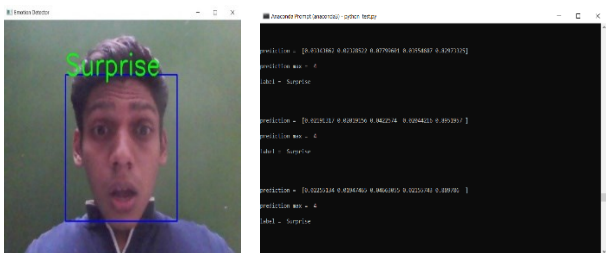

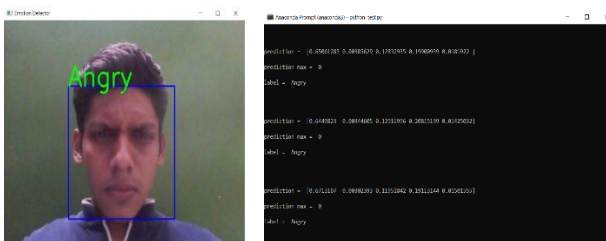
Fig 7. Prediction of Surprise Emotion
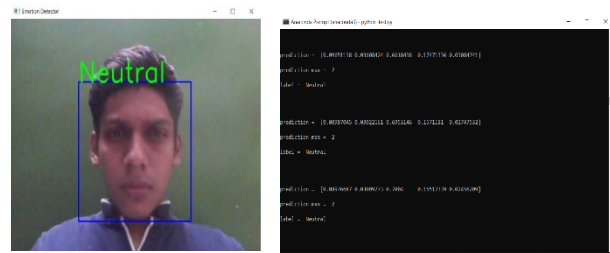

Fig 8. Prediction of Angry Emotion

Fig 9. Prediction of Neutral Emotion

Table II Comparison with other DL Models

Algorithm	Accuracy (%)	Computational Complexity
Alexnet	84	O(4)
VGG	81	O(9)
Resnet	88	0(16)
Proposed CNN	94.27	O(4)

5. Conclusion

Human facial expressions are playing important role in effective communication as well as to understand state of the mind. In recent years, many intelligent human computer interaction systems are trying to recognize human emotions through facial expressions and make effective use of it in various applications. The proposed convolutional neural network model is giving good results by recognizing emotions from real time video of human with dynamically changing emotions. Currently, the system is able to give correct predictions for all classes except disgust emotion. We are trying to add more images for all emotions in the existing pool of images to get correct recognition of all emotions from static as well as dynamic facial expressions.

References

[1] Matsugu, Masakazu, et al. "Subject independent facial expression recognition with robust face detection using a convolutional neural network." Neural Networks 16.5-6 (2003): 555-559.

[2] Zhang, Ligang, and Dian Tjondronegoro. "Facial expression recognition using facial movement features." IEEE transactions on affective computing 2.4 (2011): 219-229.

[3] Hayat, Munawar, and Mohammed Bennamoun. "An automatic framework for textured 3D video-based facial expression recognition." IEEE

Transactions on Affective Computing 5.3 (2014): 301-313.

[4] Hablani, Ramchand, Narendra Chaudhari, and Sanjay Tanwani. "Recognition of facial expressions using local binary patterns of important facial parts." International Journal of Image Processing (IJIP) 7.2 (2013): 163-170.

[5] Li, Zisheng, Jun-ichi Imai, and Masahide Kaneko. "Facial-component-based bag of words and phog descriptor for facial expression recognition." 2009 IEEE International Conference on Systems, Man and Cybernetics. IEEE, 2009.

[6] Yu, Zhiding, and Cha Zhang. "Image based static facial expression recognition with multiple deep network learning." Proceedings of the 2015 ACM on international conference on multimodal interaction. 2015.

[7] Sekaran, Sarmela AP Raja, Chin Poo Lee, and Kian Ming Lim. "Facial emotion recognition using transfer learning of AlexNet." 2021 9th International Conference on Information and Communication Technology (ICoICT). IEEE, 2021.

[8] Hussain, Shaik Asif, and Ahlam Salim Abdallah Al Balushi. "A real time face emotion classification and recognition using deep learning model." Journal of Physics: Conference Series. Vol. 1432. No. 1. IOP Publishing, 2020.

[9] Verma, Monu, Santosh Kumar Vipparthi, Girdhari Singh, and Subrahmanyam Murala. "LEARNet: Dynamic imaging network for micro expression recognition." IEEE Transactions on Image Processing 29 (2019): 1618-1627.

[10] https://www.kaggle.com/msambare/fer2013